

University of Groningen

On Association Measures for Continuous Variables and Correction for Chance

Warrens, Matthijs J.

Published in:
Journal of Probability and Statistics

DOI:
[10.1155/2015/375491](https://doi.org/10.1155/2015/375491)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Warrens, M. J. (2015). On Association Measures for Continuous Variables and Correction for Chance. *Journal of Probability and Statistics*, 2015, [375491]. <https://doi.org/10.1155/2015/375491>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Research Article

On Association Measures for Continuous Variables and Correction for Chance

Matthijs J. Warrens

GION, University of Groningen, Grote Rozenstraat 3, 9712 TG Groningen, Netherlands

Correspondence should be addressed to Matthijs J. Warrens; m.j.warrens@rug.nl

Received 24 June 2015; Revised 20 October 2015; Accepted 28 October 2015

Academic Editor: Heinz Holling

Copyright © 2015 Matthijs J. Warrens. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper studies correction for chance for association measures for continuous variables. The set of linear transformations of Pearson's product-moment correlation is used as the domain of the correction for chance function. Examples of measures in this set are Tucker's congruence coefficient, Jobson's coefficient, and Pearson's correlation. An equivalence relation on the set of linear transformations is defined. The fixed points of the correction for chance function are characterized. It is shown that each linear transformation is mapped to the fixed point in its equivalence class.

1. Introduction

In various subfields of statistics, association measures are commonly used to express the strength of a relationship between two variables in a number. Individual measures are used for summarizing parts of a research study, while matrices of association measures can be used as input for multivariate data analysis techniques like regression analysis and component analysis [1, 2]. In this paper, we study measures that quantify the association between two continuous variables. Examples of these measures are Tucker's congruence coefficient, Jobson's coefficient, and Pearson's product-moment correlation. The latter coefficient is commonly used as a measure of the linear dependence between two continuous variables.

Pearson's correlation is a measure that has multiple interesting interpretations [3]. For example, the squared correlation may be interpreted as the percentage of variance that the two variables have in common. However, for many other association measures, only the extreme values have a clear interpretation. Many association measures have maximum value unity, indicating perfect association. However, a zero value is usually attained in different situations. For example, Pearson's correlation and Jobson's coefficient have zero value under statistical independence, while Tucker's congruence coefficient has zero value if the inner product of the two variables equals zero.

It may be desirable that an association measure has zero value if two variables are statistically independent [4–6]. If a measure does not have zero value under statistical independence, a standard approach in the literature is to correct it for association due to chance [7–9]. After correction for chance, a measure A has a form

$$c(A) = \frac{A - E(A)}{1 - E(A)}, \quad (1)$$

where $E(A)$ is the value of measure A under chance and 1 in the denominator of (1) is the maximum value of A . In this paper, we only consider measures with maximum value unity. Furthermore, we assume that $E(A) < 1$ to avoid the indeterminate case $E(A) = 1$. Transformation (1) sets the value of a measure at zero under independence, while leaving the maximum value at unity.

Transformation (1) has been almost exclusively studied in relation to association measures for binary variables. Measures for binary variables are also called coefficients for 2×2 tables. Several authors showed that certain 2×2 measures become equivalent after correction (1). For example, Fleiss [8], Zegers [6], Albatineh et al. [7], and Warrens [9] showed that measures by Dice, Hamann, and Rand become Cohen's kappa after correction for chance. Furthermore, Warrens [10] characterized all 2×2 measures that are transformed into weighted kappa [10, 11] by formula (1).

The abovementioned results provide new interpretations of various chance-corrected measures. Furthermore, if different measures coincide after correction for chance, the task of choosing the best chance-corrected measure for the data-analytic task at hand becomes easier, since there are less measures to choose from. Moreover, Warrens [9, 10, 12] studied transformation (1) as a mathematical function. These latter studies revealed how different association measures from the literature are related and also provided new ways to interpret chance-corrected measures.

In this paper, we study correction for chance for association measures for continuous variables. Several results for 2×2 measures are generalized to the more complex case of continuous variables. Furthermore, various new results illustrate how certain measures are related and provide new ways for interpreting them. For example, we show that Pearson's correlation may be interpreted as a chance-corrected measure, which is a new result.

The paper is organized as follows. In Section 2, we introduce the notation and definitions of the association measures, and we specify the domain of the correction for chance function. In Section 3, we briefly discuss statistical inference based on measures in this domain. In Section 4, we define the correction for chance function and present some of its properties. Some of the results presented in this section generalize properties presented in Warrens [9] for measures for 2×2 tables. In Section 5, we consider several families of association measures that belong to the same equivalence classes. In Section 6, we characterize the fixed points of the function c in (1). Finally, Section 7 contains a discussion.

2. A Set of Association Measures

Since many association measures in the literature have been proposed as sample statistics [13, 14], we will study the measures here as sample statistics. For certain measures, the population equivalents can be obtained by replacing sample variances and covariances by their corresponding population parameters, degenerate cases excluded.

Suppose we have scores of n objects (subjects, individuals) on two variables x and y . An individual score is denoted by x_i for $i \in \{1, 2, \dots, n\}$. Let the summation \sum be short for $\sum_{i=1}^n$. The mean score of x is given by

$$\bar{x} = \frac{1}{n} \sum x_i, \quad (2)$$

the unbiased sample variance is given by

$$s_x^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2, \quad (3)$$

and the mean squared value of x is given by

$$t_x^2 = \frac{1}{n} \sum x_i^2. \quad (4)$$

The sample statistics for y are defined analogously. The unbiased sample covariance of x and y is given by

$$s_{xy} = \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y}). \quad (5)$$

Association measures can be defined as functions from the set of pairs of variables of length n into the real numbers. The set

$$D = \{(x, y) \mid x_i, y_i \in \mathbb{R}\} \quad (6)$$

will be used as the domain. For many measures from the literature, the codomain is either the interval $[0, 1]$ or the interval $[-1, 1]$. An association measure is defined as a function $A : D \rightarrow \mathbb{R}$, and the set of all such measures is denoted by $K = \{A : D \rightarrow \mathbb{R}\}$. A measure that belongs to K is Pearson's correlation coefficient given by

$$r(x, y) = \frac{s_{xy}}{s_x s_y} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{(n-1) s_x s_y}. \quad (7)$$

Pearson's correlation is a commonly used measure of linear dependence between two variables.

In the following, we are interested in a particular subset of K . Kendall and Stuart [15, p. 492] discuss the permutation distribution of r . The exact distribution of r can be obtained by considering all $n!$ permutations of the scores on x , while keeping the scores on y fixed. Under permutation of the scores on x , only $\sum x_i y_i$ is a random variable. For each permutation, we may calculate $\sum x_i y_i$. The expectation $E(\sum x_i y_i)$ can then be defined as the average of all the values of $\sum x_i y_i$. Kendall and Stuart [15, p. 492] show that

$$E\left(\sum x_i y_i\right) = n \bar{x} \bar{y}. \quad (8)$$

Since only $\sum x_i y_i$ is a random variable under permutation, we can write r as a linear transformation of $\sum x_i y_i$; that is, $\lambda + \mu \sum x_i y_i$, where

$$\begin{aligned} \lambda &= \frac{-n \bar{x} \bar{y}}{(n-1) s_x s_y}, \\ \mu &= \frac{1}{(n-1) s_x s_y} \end{aligned} \quad (9)$$

are functions that are not affected by permutation of the scores on x or y . More generally, we may consider the set $L \subset K$ given by

$$L = \left\{A : D \rightarrow \mathbb{R} \mid A = \lambda + \mu \sum x_i y_i, A \leq 1\right\}. \quad (10)$$

The elements of L are linear transformations of $\sum x_i y_i$, and $\lambda = \lambda(x, y)$ and $\mu = \mu(x, y)$ are functions that remain unchanged under permutations of the scores on x or y . Furthermore, 1 is the maximum value of the measures in L . Since Pearson's correlation r belongs to L , the set is nonempty. Examples of other association measures that belong to L are presented in Section 5.

Instead of considering L , we may also consider the set $M \subset K$ given by

$$M = \left\{A : D \rightarrow \mathbb{R} \mid A = \lambda + \mu r, A \leq 1\right\}. \quad (11)$$

Measures in M are linear transformations of r , and λ and μ are functions that are unaffected by permutations of the scores on x or y . Lemma 1 shows that L and M are equivalent.

Lemma 1. $L = M$.

Proof. Consider (7). We can write $\sum x_i y_i$ as $\lambda + \mu r$, where

$$\begin{aligned}\lambda &= n\bar{x}\bar{y}, \\ \mu &= (n-1)s_x s_y.\end{aligned}\quad (12)$$

Hence, $\sum x_i y_i \in M$ and it follows that $L \subset M$. Furthermore, we can write r as $\lambda + \mu \sum x_i y_i$, where λ and μ are presented in (9). Hence, $r \in L$ and it follows that $M \subset L$. Thus, $L = M$. \square

Lemma 1 shows that linearity in $\sum x_i y_i$ is equivalent to linearity in r . The proof of Lemma 1 illustrates that any element of L can be used as a generator of L . For example, we have $L = \langle r \rangle$. The set L is used as the domain of the correction for chance formula in the next section. A similar family of linear transformations has been considered in [5, 7, 9, 10, 16].

3. Statistical Inference

In this paper, we mainly present various algebraic and analytic properties of the broad family of association measures in (10) and (11). In this section, we briefly discuss statistical inference based on members of this family. In the framework presented here, variables are considered random if they are not invariant under permutation [14, 17]. Therefore, the quantity $\sum x_i y_i$ is the only random variable in (10). A straightforward approach to performing hypothesis tests or constructing confidence intervals for a member of the family in (10) could be based on a permutation test with respect to the quantity $\sum x_i y_i$. Furthermore, confidence intervals for a member of the family in (10) may be obtained using the bootstrap.

It should be noted that the assumption that variables are random only if they are not invariant under permutation is unusual for Pearson's correlation coefficient. As pointed out by a reviewer, with Pearson's correlation, it is common to assume that x and y follow a bivariate normal distribution. In this case, the quantities \bar{x} , \bar{y} , s_x , s_y , t_x , and t_y are all random variables.

4. Correction for Chance

Equation (1) presents the formula of a measure A after correction for chance. For measures in L , the correction for chance function is defined as

$$\begin{aligned}c : L &\longrightarrow L, \\ A &\longmapsto \frac{A - E(A)}{1 - E(A)}.\end{aligned}\quad (13)$$

Lemma 2 shows that, for measures in L , formula (13) becomes formula (14).

Lemma 2. Let $A \in L$ with $A = \lambda + \mu \sum x_i y_i$. One has

$$c(A) = \frac{(n-1)s_{xy}}{(1-\lambda)/\mu - n\bar{x}\bar{y}}. \quad (14)$$

Proof. Using identity (8), we have for measures in L the identity

$$\begin{aligned}E(A) &= E\left(\lambda + \mu \sum x_i y_i\right) = \lambda + \mu E\left(\sum x_i y_i\right) \\ &= \lambda + \mu n\bar{x}\bar{y}.\end{aligned}\quad (15)$$

Using identity (15) in formula (13) and dividing all terms in the result by μ , we obtain

$$c(A) = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(1-\lambda)/\mu - n\bar{x}\bar{y}}, \quad (16)$$

or, equivalently, formula (14). \square

Formula (14) shows that elements of L coincide after correction for chance if they have the same ratio

$$\frac{1-\lambda}{\mu}. \quad (17)$$

This property of elements in L will be used repeatedly in the following. A similar property was found in [7] for validation indices in cluster analysis. Furthermore, for association measures for 2×2 tables, a similar property was found in [9].

The function c in (13) is a map from L to L if L is closed under c . Lemma 3 shows that this is the case.

Lemma 3. L is closed under c .

Proof. Let $A \in L$ with $A = \lambda + \mu \sum x_i y_i$. The formula for $c(A)$ is presented in (16). We can write $c(A)$ as $\lambda^* + \mu^* \sum x_i y_i$, where

$$\begin{aligned}\lambda^* &= \frac{-n\bar{x}\bar{y}}{(1-\lambda)/\mu - n\bar{x}\bar{y}}, \\ \mu^* &= \frac{1}{(1-\lambda)/\mu - n\bar{x}\bar{y}}.\end{aligned}\quad (18)$$

Hence, $c(A) \in L$. \square

Let $A \in L$, let $a, b \in \mathbb{R}$ with $b \neq 0$, and consider the linear transformation $B = a + bA$. Since we require that all association measures have maximum value 1, we have the restriction $a + b = 1$. Hence, $B = 1 - b + bA$ is also an element in L . Lemma 4 shows that A and B coincide after correction for chance. A proof of Lemma 4 for the special case of 2×2 measures can be found in Warrens [9].

Lemma 4. Let $A \in L$, let $b \in \mathbb{R}$, and consider $B = 1 - b + bA$. One has $c(A) = c(B)$.

Proof. Since $A \in L$, there exist λ and μ such that $A = \lambda + \mu \sum x_i y_i$. Hence, $B = 1 - b + b(\lambda + \mu \sum x_i y_i)$ and we can write B as $\lambda^* + \mu^* \sum x_i y_i$, where

$$\begin{aligned}\lambda^* &= 1 - b + b\lambda, \\ \mu^* &= b\mu.\end{aligned}\quad (19)$$

Using these quantities, ratio (17) is equal to

$$\frac{1-\lambda^*}{\mu^*} = \frac{b(1-\lambda)}{b\mu} = \frac{1-\lambda}{\mu}. \quad (20)$$

Since ratio (17) is identical for A and B , it follows that $c(A) = c(B)$. \square

Suppose we have several association measures that coincide after correction for chance. Lemma 5 shows that the arithmetic mean of the association measures becomes identical to the measures after correction for chance. A proof of Lemma 5 for the special case of 2×2 measures can be found in Warrens [9].

Lemma 5. Let $A_\ell \in L$ with $A_\ell = \lambda_\ell + \mu_\ell \sum x_i y_i$ for $\ell \in \{1, 2, \dots, m\}$ such that

$$\frac{1 - \lambda_1}{\mu_1} = \frac{1 - \lambda_2}{\mu_2} = \dots = \frac{1 - \lambda_m}{\mu_m}. \quad (21)$$

The arithmetic mean $B = m^{-1} \sum_{\ell=1}^m A_\ell$ coincides with A_ℓ after correction for chance.

Proof. We can write B as $\lambda^* + \mu^* \sum x_i y_i$, where

$$\begin{aligned} \lambda^* &= \frac{1}{m} \sum_{\ell=1}^m \lambda_\ell, \\ \mu^* &= \frac{1}{m} \sum_{\ell=1}^m \mu_\ell. \end{aligned} \quad (22)$$

Using these quantities, ratio (17) is equal to

$$\frac{1 - \lambda^*}{\mu^*} = \frac{m - \sum_{\ell=1}^m \lambda_\ell}{\sum_{\ell=1}^m \mu_\ell} = \frac{\sum_{\ell=1}^m \mu_\ell ((1 - \lambda_\ell) / \mu_\ell)}{\sum_{\ell=1}^m \mu_\ell}. \quad (23)$$

The right-hand side of this equation shows that $(1 - \lambda^*) / \mu^*$ is a weighted average of $(1 - \lambda_\ell) / \mu_\ell$ using weights μ_ℓ . Since $(1 - \lambda_\ell) / \mu_\ell$ are all identical, $(1 - \lambda^*) / \mu^*$ is identical to $(1 - \lambda_\ell) / \mu_\ell$. Hence, $c(A_\ell) = c(B)$. \square

5. Equivalence Classes

Ratio (17) divides the association measures in L into equivalence classes with respect to function c in (14). We have the following definition. Two measures $A, B \in L$ are said to be equivalent with respect to (14), denoted by $A \sim B$, if they have the same ratio (17). It can be shown that \sim is an equivalence relation on L ; that is, \sim is reflexive, symmetric, and transitive.

The equivalence relation \sim divides the elements of L into equivalence classes, one class for each value of (17). We consider some examples. Let $\nu = \nu(x, y)$ be a function that remains unchanged under permutations of the scores on x or y , and consider the family

$$\frac{2 \sum x_i y_i + \nu}{n(t_x^2 + t_y^2) + \nu}. \quad (24)$$

If $\nu = 0$, (24) is identical to the coefficient of identity [6, 14, 17]. If $\nu = -2n\bar{x}\bar{y}$, (24) is identical to Jobson's coefficient [18],

$$e(x, y) = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n/2)(t_x^2 + t_y^2) - n\bar{x}\bar{y}}, \quad (25)$$

which is identical to the chance-corrected identity coefficient [6, 17]. We can write (24) as $\lambda + \mu \sum x_i y_i$, where

$$\begin{aligned} \lambda &= \frac{\nu}{n(t_x^2 + t_y^2) + \nu}, \\ \mu &= \frac{2}{n(t_x^2 + t_y^2) + \nu}. \end{aligned} \quad (26)$$

Using λ and μ in (26), ratio (17) is equal to

$$\frac{1 - \lambda}{\mu} = \frac{n(t_x^2 + t_y^2)}{2}. \quad (27)$$

Hence, all special cases of (24), regardless of the specification of ν , belong to the equivalence class that is characterized by ratio (27). Furthermore, all special cases of (24) coincide after correction for chance.

Next, consider the family

$$\frac{\sum x_i y_i + \nu}{nt_x t_y + \nu}. \quad (28)$$

If $\nu = 0$, (28) is identical to Tucker's congruence coefficient [14, 19], also called the coefficient of proportionality [6, 14]. If $\nu = -n\bar{x}\bar{y}$, (28) is identical to the chance-corrected proportionality coefficient [6, 17]. We can write (28) as $\lambda + \mu \sum x_i y_i$, where

$$\begin{aligned} \lambda &= \frac{\nu}{nt_x t_y + \nu}, \\ \mu &= \frac{1}{nt_x t_y + \nu}. \end{aligned} \quad (29)$$

Using λ and μ in (29), ratio (17) is equal to

$$\frac{1 - \lambda}{\mu} = nt_x t_y. \quad (30)$$

Hence, all special cases of (28) belong to the equivalence class that is characterized by ratio (30). Furthermore, all special cases of (28) coincide after correction for chance.

Next, consider the family

$$\frac{\sum x_i y_i - n\bar{x}\bar{y} + \nu}{(n-1)s_x s_y + \nu}. \quad (31)$$

If $\nu = 0$, (31) is identical to Pearson's correlation in (7). We can write (31) as $\lambda + \mu \sum x_i y_i$, where

$$\begin{aligned} \lambda &= \frac{-n\bar{x}\bar{y} + \nu}{(n-1)s_x s_y + \nu}, \\ \mu &= \frac{1}{(n-1)s_x s_y + \nu}. \end{aligned} \quad (32)$$

Using λ and μ in (32), ratio (17) is equal to

$$\frac{1 - \lambda}{\mu} = (n-1)s_x s_y + n\bar{x}\bar{y}. \quad (33)$$

Hence, all special cases of (31) belong to the equivalence class that is characterized by ratio (33). Furthermore, all special cases of (31) coincide after correction for chance.

The examples in (24), (28), and (31) illustrate that many measures from the literature belong to the set L . Furthermore, the examples show that the equivalence classes have infinitely many elements. Since multiple measures coincide after correction for chance, the function c in (14) is not injective. Since c is not injective, it is not invertible. Furthermore, for $A, B \in L$, we have, in general, $c(A + B) \neq c(A) + c(B)$. Thus, in general, c is not a linear map. Lemmas 4 and 5 show that c is a linear map if the domain of c is restricted to an equivalence class for a specific value of ratio (17). All association measures in this equivalence class and their linear combinations are mapped onto a measure of the form (14). In the next section, we characterize these so-called fixed points.

6. Fixed Points

A measure $A \in L$ is called a fixed point of c if $c(A) = A$ holds. Using λ^* and μ^* in (18), we have

$$\frac{1 - \lambda^*}{\mu^*} = \frac{1 - \lambda}{\mu}. \quad (34)$$

Hence, $c(c(A)) = c(A)$. In other words, the function c is idempotent. An idempotent function has at least one fixed point. Suppose $A \in L$ is not a fixed point and let $c(A) = B$. Since c is idempotent, we have $c(B) = c(c(A)) = c(A) = B$. Hence, elements of L that are not fixed points are mapped to fixed points.

In the remainder of this section, we characterize the fixed points of c . Let $\nu = \nu(x, y)$ be a function that remains unchanged under permutations of the scores on x or y , and consider the set $F \subset L$ given by

$$F = \left\{ A \in L \mid A = \frac{s_{xy}}{\nu}, \text{ for some } \nu \neq 0 \right\}. \quad (35)$$

The elements in F are association measures of the form s_{xy}/ν , where s_{xy} is the sample covariance of variables x and y . Lemma 6 shows that F is the set of fixed points of c .

Lemma 6. *The set $F \subset L$ is the set of fixed points of c .*

Proof. (\Rightarrow) Let $A \in F$. Then,

$$A = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1)\nu}, \quad (36)$$

for some $\nu \neq 0$. Using

$$\begin{aligned} \lambda &= \frac{-n\bar{x}\bar{y}}{(n-1)\nu}, \\ \mu &= \frac{1}{(n-1)\nu} \end{aligned} \quad (37)$$

in (16), we obtain

$$c(A) = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1)\nu}. \quad (38)$$

Hence, $c(A) = A$, and it follows that A is a fixed point.

(\Leftarrow) Let $A \in L$ with $A = \lambda + \mu \sum x_i y_i$ be a fixed point. Then, $A = c(A)$, or equivalently

$$\lambda + \mu \sum x_i y_i = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(1-\lambda)/\mu - n\bar{x}\bar{y}}. \quad (39)$$

Equating $\sum x_i y_i$ -parts and “not”- $\sum x_i y_i$ -parts on both sides of the equality, we obtain the identities

$$\begin{aligned} \lambda &= \frac{-n\bar{x}\bar{y}}{(1-\lambda)/\mu - n\bar{x}\bar{y}}, \\ \mu &= \frac{1}{(1-\lambda)/\mu - n\bar{x}\bar{y}}. \end{aligned} \quad (40)$$

Using the identity $(n-1)\nu = (1-\lambda)/\mu - n\bar{x}\bar{y}$, we can write A as

$$A = \lambda + \mu \sum x_i y_i = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1)\nu} = \frac{s_{xy}}{\nu}. \quad (41)$$

Hence, $A \in F$. This completes the proof. \square

Since points that are not fixed points of c are mapped to fixed points, we have $c(L) = F$; that is, the images of c are the fixed points in L . It follows from Lemma 6 that Pearson's correlation r in (7) and Jobson's coefficient e in (25) are two fixed points of c . All measures that belong to the equivalence class characterized by ratio (33) are mapped to Pearson's correlation. All measures that belong to the equivalence class characterized by ratio (27) are mapped to Jobson's coefficient.

7. Discussion

In this paper, we studied correction for chance as a mathematical function in the context of association measures for continuous variables. The function is not injective and in general not a linear map. All association measures that are not fixed points of the function are mapped to fixed points. Notable fixed points are Jobson's coefficient and Pearson's correlation. Both coefficients are fixed points of distinct equivalence classes of association measures. Characteristics of association measures in both classes were presented.

The results show that both Jobson's coefficient and Pearson's correlation may be interpreted as chance-corrected measures. For Pearson's correlation, this interpretation is new. Thirteen other ways to look at Pearson's correlation are discussed in Rodgers and Nicewander [3]. What is also new is that there are infinitely many measures that become Jobson's coefficient and infinitely many coefficients that coincide with Pearson's correlation after correction for chance. A data-analytic choice between Jobson's coefficient and Pearson's correlation could be based on the metric scales discussed in Zegers [6, 17].

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The author thanks Professor Heinz Holling for the helpful comments and valuable suggestions on an earlier version of this paper.

References

- [1] J. C. Gower, "Some distance properties of latent root and vector methods used in multivariate analysis," *Biometrika*, vol. 53, pp. 325–338, 1966.
- [2] J. C. Gower and P. Legendre, "Metric and Euclidean properties of dissimilarity coefficients," *Journal of Classification*, vol. 3, no. 1, pp. 5–48, 1986.
- [3] J. L. Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," *The American Statistician*, vol. 42, no. 1, pp. 59–66, 1988.
- [4] R. Popping, *Overeenstemmingsmaten voor nominale data*, Rijksuniversiteit Groningen, Groningen, The Netherlands, 1983.
- [5] M. J. Warrens, "On association coefficients for 2×2 tables and properties that do not depend on the marginal distributions," *Psychometrika*, vol. 73, no. 4, pp. 777–789, 2008.
- [6] F. E. Zegers, *A General Family of Association Coefficients*, Boomker, Groningen, The Netherlands, 1986.
- [7] A. N. Albatineh, M. Niewiadomska-Bugaj, and D. Mihalko, "On similarity indices and correction for chance agreement," *Journal of Classification*, vol. 23, no. 2, pp. 301–313, 2006.
- [8] J. L. Fleiss, "Measuring agreement between two judges on the presence or absence of a trait," *Biometrics. Journal of the Biometric Society*, vol. 31, no. 3, pp. 651–659, 1975.
- [9] M. J. Warrens, "On similarity coefficients for 2×2 tables and correction for chance," *Psychometrika*, vol. 73, no. 3, pp. 487–502, 2008.
- [10] M. J. Warrens, "Chance-corrected measures for 2×2 tables that coincide with weighted kappa," *British Journal of Mathematical and Statistical Psychology*, vol. 64, no. 2, pp. 355–365, 2011.
- [11] M. J. Warrens, "Some paradoxical results for the quadratically weighted kappa," *Psychometrika*, vol. 77, no. 2, pp. 315–323, 2012.
- [12] M. J. Warrens, "On association coefficients, correction for chance, and correction for maximum value," *Journal of Modern Mathematics Frontier*, vol. 2, no. 4, pp. 111–119, 2013.
- [13] M. J. Warrens, "Corrected Zegers-ten Berge coefficients are special cases of Cohen's weighted kappa," *Journal of Classification*, vol. 31, no. 2, pp. 179–193, 2014.
- [14] F. E. Zegers and J. M. F. ten Berge, "A family of association coefficients for metric scales," *Psychometrika*, vol. 50, no. 1, pp. 17–24, 1985.
- [15] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics, Volume 2: Inference and Relationship*, Charles Griffin, London, UK, 1973.
- [16] A. N. Albatineh, "Means and variances for a family of similarity indices used in cluster analysis," *Journal of Statistical Planning and Inference*, vol. 140, no. 10, pp. 2828–2838, 2010.
- [17] F. E. Zegers, "A family of chance-corrected association coefficients for metric scales," *Psychometrika*, vol. 51, no. 4, pp. 559–562, 1986.
- [18] J. D. Jobson, "A coefficient of equality for questionnaire items with interval scales," *Educational and Psychological Measurement*, vol. 36, no. 2, pp. 271–274, 1976.
- [19] U. Lorenzo-Seva and J. M. F. ten Berge, "Tucker's congruence coefficient as a meaningful index of factor similarity," *Methodology*, vol. 2, no. 2, pp. 57–64, 2006.

